Научная статья
УДК 81`33

Цифровой инструментарий автоматизированной обработки текста и речи в условиях перехода на свободные операционные системы семейства Linux

М. В. Каменский

Северо-Кавказский федеральный университет, Ставрополь, Россия
stavdev@mail.ru

Аннотация. Целью настоящей работы является систематизация современного цифрового инструментария автоматизированной обработки текста и речи в контексте его совместимости с операционными системами семейства Linux. Материалом исследования послужили каталоги и реестры программного обеспечения общего и лингвистического назначения, которое разрабатывается международными и российскими коллективами. В результате анализа, проведенного с применением описательного и экспериментального методов, уточнена функциональность лингвистического программного обеспечения в условиях его работы под управлением операционной системы семейства Linux. Предложена модель повышения эффективности рабочего процесса лингвиста-исследователя в условиях использования отечественных операционных систем исследованного семейства.

Ключевые слова: цифровая лингвистика, компьютерная лингвистика, автоматизированная обработка текста, цифровой инструментарий, свободное программное обеспечение, операционная система Linux

Для цитирования: Каменский М. В. Цифровой инструментарий автоматизированной обработки текста и речи в условиях перехода на операционные системы семейства Linux // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2025. Вып. 9 (903). С. 38–44.

Original article

Digital Toolkit for Natural Language Processing in the Linux Free Operating System Family

Mikhail V. Kamensky

North-Caucasus Federal University, Stavropol, Russia
stavdev@mail.ru

Abstract. The aim of the present study is to determine the compatibility status of the contemporary digital toolkit for natural language processing in the Linux kernel-based operating system environment. The research is based on the material of catalogs and registries of general purpose and linguistic software developed by international and Russian programmer teams. The analysis, conducted by means of the descriptive and experimental methods, describes the functionality of specialized linguistic software within the framework of their use in the Linux operating system environment. A model of improving the efficiency of the Linux-centric linguistics digital workflow is proposed, built around the use of Russian operating systems based on the Linux kernel.

Keywords: digital linguistics, computational linguistics, natural language processing, digital toolkit, free software, Linux operating system

For citation: Kamensky, M. V. (2025). Digital toolkit for natural language processing in the Linux free operating system family. Vestnik of Moscow State Linguistic University. Humanities, 9(903), 38–44. (In Russ.)

ВВЕДЕНИЕ

В настоящей работе ставится и решается проблема обеспечения эффективности научного поиска в сфере теоретической и прикладной лингвистики. Рассматривается применение цифрового инструментария в современных условиях интеграции импортозамещающих технологий и отечественного программного обеспечения.

Цель исследования – охарактеризовать функциональность и совместимость основных программных продуктов общего и профессионального назначения, систематически используемых при проведении автоматизированной обработки текста в лингвистических исследованиях. Инструментарий этой работы составляют свободные операционные системы семейства Linux в целом и их отечественные реализации в частности.

Актуальность исследования вызвана необходимостью повышения эффективности автоматизации и алгоритмизации исследовательских процедур в сфере теоретической и прикладной лингвистики в изменяющихся технологических условиях, подразумевающих качественный переход от проприетарных решений (ОС Microsoft Windows и совместимое проприетарное ПО общего назначения) к свободным и открытым решениям с преимущественным применением отечественного ПО (ОС Astra Linux, ALT Linux, РЕД ОС и совместимое свободное ПО общего назначения).

Для реализации поставленной цели в пределах настоящей статьи решаются следующие задачи:

- предпринимается аналитический обзор современных научных исследований по проблеме цифровизации научно-исследовательского рабочего процесса лингвиста;
- проводится систематизация существующего ПО общего и профессионального назначения, применимого для автоматизированной обработки текста и звучащей речи и совместимого с ОС на базе ядра Linux;
- осуществляется разработка рекомендаций по повышению эффективности организации цифрового рабочего процесса лингвиста в условиях перехода на отечественные ОС семейства Linux.

Научная новизна предлагаемой работы состоит в следующем:

- в рамках единого цифрового рабочего процесса систематизировано актуальное лингвистическое программное обеспечение, совместимое с операционными системами семейства Linux;
- получены новые данные о функциональном потенциале ПО общего и профессионального назначения для решения задач автоматизированной обработки языка и речи;
- оценены возможности повышения эффективности цифрового рабочего процесса лингвиста в условиях реализации импортозамещающих практик по интеграции отечественного программного обеспечения.

АНАЛИТИЧЕСКИЙ ОБЗОР ПО ПРОБЛЕМЕ ЦИФРОВИЗАЦИИ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОГО РАБОЧЕГО ПРОЦЕССА ЛИНГВИСТА

В последние годы в связи со всё более активной интеграцией цифровых технологий в различные сферы профессиональной деятельности и развитием новых цифровых решений для различных задач, в том числе технологий искусственного интеллекта, а также в связи с реализацией государственной политики импортозамещения и поэтапного перехода на отечественное программное обеспечение все более актуальным становится вопрос о поддержании и повышении эффективности решения практических и научно-исследовательских задач в новых технологических условиях.

Применительно к научно-исследовательской деятельности в области лингвистики данный процесс диктует, в первую очередь, необходимость каталогизации и систематизации профессиональных ресурсов и программного обеспечения лингвистического назначения, что находит отражение в современных научных трудах отечественных ученых [Антопольский, 2021; Каменский, 2021].

Параллельно данным научным изысканиям также активно развиваются отечественные лингвистические программные продукты, в частности, отечественные электронные корпусы текстов, текстовые и аудиальные базы данных, а также программные алгоритмы для их обработки и анализа. Такого рода программное обеспечение разрабатывается как отдельными инициативными исследователями, так и научными коллективами [Горожанов, 2024; Гуртуева, 2024; Gorozhanov, Guseynova, 2020].

В наше время наблюдается активное развитие технологий искусственного интеллекта. Оно влечет за собой развитие программных решений в сфере автоматического перевода, генерации текста и синтеза речи, классификации документов и автоматизированного анализа текстов на основе машинного обучения. Существенное исследовательское внимание направлено, с одной стороны, на общую реализацию функционального потенциала искусственного интеллекта в гуманитарной сфере [Аветисян, 2024], а с другой – на практическое

применение больших языковых моделей в лингвистике и лингводидактике [Авраменко, 2023].

Динамично развивается общее интегративное направление цифровой лингвистики, объединяющее концепции компьютерной, корпусной, квантитативной, математической лингвистики, а также ряд смежных областей, связанных с цифровизацией различных исследовательских траекторий в сфере филологического знания и прикладных областей лингвистической деятельности [Поликарпов, 2019]. В научных трудах исследуется такое понятие, как «цифровая филология», изучается текущее состояние и обсуждаются перспективы развития цифрового направления гуманитарных наук [Шейко, 2023].

Множество научно-исследовательских работ посвящено собственно вопросам автоматизированной обработки текста. Так, в последние годы внимание исследователей было направлено на сравнение результатов автоматизированной и ручной обработки текста при решении задачи сентимент-анализа [Гималетдинова, Довтаева, 2021], решались вопросы автоматизации получения лингвостатистических данных при системной обработке текста [Максименко, 2019], разрабатывались принципы автоматизации анализа ритма текста [Бойчук, 2021].

Однако при активном развитии цифровой лингвистики и наличии множества работ, посвященных вопросам автоматизированной обработки текста и речи, в означенной сфере знания всё еще имеются существенные пробелы. Вопросы цифровизации рабочего процесса лингвиста в условиях перехода на свободное программное обеспечение, на операционные системы семейства Linux, а также на отечественные программные продукты достаточно редко оказываются в центре исследовательского внимания [Фаткулин, 2015].

Считаем, что активный курс на внедрение импортозамещающего программного обеспечения, а также активизация процессов перехода на ОС Linux в высших учебных заведениях и научно-исследовательских институтах в полной мере актуализируют вопрос специального рассмотрения функционального потенциала и совместимости существующих лингвистических программных продуктов с отечественными операционными системами на основе ядра Linux.

МОДЕЛЬ ЦИФРОВОГО РАБОЧЕГО ПРОЦЕССА ЛИНГВИСТА В УСЛОВИЯХ ПЕРЕХОДА НА ОПЕРАЦИОННЫЕ СИСТЕМЫ СЕМЕЙСТВА LINUX

Предпринятое исследование существующих дистрибутивов операционных систем на базе ядра Linux показало, что в высших учебных заведениях и научных организациях России в настоящее время в рамках перехода на отечественное программное обеспечение[1] используются три операционные системы данного семейства: 1) ALT Linux (Альт СП)[2]; 2) Astra Linux[3]; 3) РЕД ОС[4]. Они относятся к системам общего назначения, сертифицированы ФСТЭК России и включены в единый реестр российских программ для ЭВМ и баз данных[5].

В функциональном отношении данные дистрибутивы в существенной мере схожи, все они совместимы с существующим открытым и свободным программным обеспечением общего и профессионального лингвистического назначения. Данное сходство обеспечивается использованием современной версии ядра Linux (6.x) и стандартных библиотек времени выполнения для 64-битной архитектуры ПК x86-64, а также наличием развитых репозиториев с пакетной базой. Она включает все необходимые компоненты для запуска и компиляции существующих приложений для ОС Linux. Так, репозиторий Astra Linux основан на пакетной базе Debian GNU/Linux формата DEB, репозитории РЕД ОС и Альт СП – на пакетной базе RPM (в случае Альт СП основой выступила пакетная база Red Hat Enterprise Linux). Используемый по умолчанию интерфейс рабочего стола пользователя в трех исследованных операционных системах различен, однако во всех трех случаях имеет место классическая парадигма оконного интерфейса. Она призвана быть максимально прозрачной и понятной для пользователей, долгое время работающих в оконной среде ОС Microsoft Windows.

Обзор пакетной базы российских дистрибутивов Linux показал, что к основному программному обеспечению общего назначения, представленному в репозиториях либо загружаемому из совместимых сторонних источников и способному выступить аналогом и альтернативой существующему распространенному проприетарному программному обеспечению, типичному для персональных компьютеров, работающих под управлением ОС Microsoft Windows, следует причислить следующее ПО:

1. LibreOffice[6] – офисный пакет, включающий в себя текстовый процессор Writer, редактор электронных таблиц Calc, редактор презентаций Impress, базу данных Base

[1]URL: https://minobrnauki.gov.ru/importozameshcheniye/ (дата обращения: 19.05.2025).
[2]URL: https://www.basealt.ru/ (дата обращения: 19.05.2025).
[3]URL: https://astralinux.ru/ (дата обращения: 19.05.2025).
[4]URL: https://redos.red-soft.ru/ (дата обращения: 19.05.2025).
[5]URL: https://reestr.digital.gov.ru/ (дата обращения: 20.05.2025).
[6]URL: https://www.libreoffice.org/ (дата обращения: 20.05.2025).

и другое ПО, в актуальной версии обладающее высокой степенью совместимости с пакетом Microsoft Office[1].

2. Chromium[2] – интернет-браузер, основанный на одноименном ядре, также лежащем в основе многих распространенных браузеров (Chrome, Edge, Яндекс Браузер и др.).
3. Thunderbird[3] – почтовый клиент для работы с электронной почтой и группами новостей.
4. VLC Media Player[4] – медиапроигрыватель с интегрированными аудио- и видеокодеками, позволяющий проигрывать существующий медиаконтент в подавляющем большинстве существующих форматов[5].
5. Okular[6] – средство просмотра текстовых документов в различных распространенных форматах (pdf, djvu, epub и др.).

Также подчеркнем, что в рамках реализации политики импортозамещения создан отдельный каталог российского программного обеспечения. Данный каталог систематизирует отечественное ПО, выступающее прямой альтернативой иностранному проприетарному ПО[7]. В данном перечне ПО общего назначения, релевантное для сопровождения лингвистической деятельности, приведено в каталогах «ПО для работы с документами и текстами»[8], «Офисные пакеты»[9], «Коммуникационное ПО»[10], «Аудио, видео, обработка изображений»[11]. Большая часть представленного в перечне ПО является коммерческим проприетарным программным обеспечением. Поэтому необходимо признать существование в настоящий момент двух путей альтернативизации программного обеспечения в рамках цифрового процесса лингвиста:

1) свободное и открытое программное обеспечение, разрабатываемое международными коллективами программистов (в том числе с российским участием) и распространяемое в большинстве случаев по некоммерческим свободным лицензиям;
2) отечественное проприетарное программное обеспечение, разрабатываемое российскими коллективами программистов и распространяемое в большинстве случаев на коммерческой основе по корпоративным лицензиям.

В обоих случаях наблюдается высокая совместимость ПО общего назначения с операционными системами семейства Linux, в силу чего решение об использовании каждого конкретного подхода (либо об их совмещении) может приниматься на индивидуальной основе в каждой конкретной организации или в определенном научном коллективе.

Применительно к собственно лингвистическому программному обеспечению отметим, что результаты исследования также показывают высокую степень совместимости специализированных программных продуктов с операционными системами на базе ядра Linux. Это объясняется академической природой такого ПО, его систематической разработкой в научных коллективах высших учебных заведений по открытой и свободной модели распространения научных достижений и технических средств получения научных данных, что максимально приоритизирует совместимость данных программных продуктов с открытыми и свободными операционными системами.

К основному лингвистическому ПО, относимому к свободному программному обеспечению с открытым кодом, совместимому с операционными системами семейства Linux и используемому для решения задач в области автоматизированной обработки текста и речи, следует причислить следующее:

1. Корпусные менеджеры GATE[12] и LancsBox[13]. GATE представляет собой разветвленную модульную инфраструктуру для разработки, сопровождения и практического применения разноязычных и разнонаправленных электронных корпусов текстов, а также для проведения автоматизированного анализа текста по различным пользовательским грамматическим и лексико-семантическим критериям на базе данных электронных корпусов. LancsBox также является корпусным менеджером для разработки электронных корпусов текстов и их практического применения и автоматизированной обработки. Однако акцент в данном программном продукте смещен в сторону реализации стандартных исследовательских процедур (KWIC, GraphColl, N-Grams и др.) и наглядной визуализации

[1]URL: https://wiki.documentfoundation.org/Feature_Comparison:_LibreOffice_-_Microsoft_Office/ru (дата обращения: 20.05.2025).
[2]URL: https://www.chromium.org/chromium-projects/ (дата обращения: 21.05.2025).
[3]URL: https://www.thunderbird.net (дата обращения: 21.05.2025).
[4]URL: https://www.videolan.org/vlc/ (дата обращения: 21.05.2025).
[5]URL: https://www.videolan.org/vlc/features.html (дата обращения: 21.05.2025).
[6]URL: https://okular.kde.org/ (дата обращения: 21.05.2025).
[7]URL: https://catalog.arppsoft.ru/replacement (дата обращения: 22.05.2025).
[8]URL: https://catalog.arppsoft.ru/replacement/section_6074055 (дата обращения: 22.05.2025).
[9]URL: https://catalog.arppsoft.ru/replacement/section_6046988 (дата обращения: 22.05.2025).
[10]URL:https://catalog.arppsoft.ru/replacement/section_6053309 (дата обращения: 22.05.2025).
[11]URL: https://catalog.arppsoft.ru/replacement/section_6050571 (дата обращения: 22.05.2025).
[12]URL: https://gate.ac.uk/ (дата обращения: 23.05.2025).
[13]URL: https://lancsbox.lancs.ac.uk/ (дата обращения: 23.05.2025).

результатов исследования с помощью встроенных механизмов.

2. Язык программирования Python[1] и специализированные лингвистические библиотеки для данного языка, реализующие алгоритмы автоматизированной обработки текста. В частности, к таким библиотекам относятся распространенные в настоящее время NLTK (Natural Language Toolkit)[2] и spaCy[3], а также библиотеки-«обертки» (от англ. wrapper) для них, предоставляющие более удобные интерфейсы разработки приложений (API) для типовых задач, например, TextBlob[4] для реализации стандартных процедур анализа англоязычных текстов и TextaCy[5] для упрощения работы со стандартными алгоритмами библиотеки spaCy. Также к данной категории ПО относятся среды разработки приложений, поддерживающие язык программирования Python, такие как Visual Studio Code[6] и ее полностью свободная реализация VS Codium[7], исключающая проприетарные компоненты и телеметрию.

3. Фонетический анализатор Praat[8], реализующий обширный функционал в области акустического анализа фонограмм, в том числе формантного анализа, спектрального анализа, анализа интонационных кривых и т.д., а также в области синтеза речи. Praat также обладает необходимым набором функций для визуализации результатов фонетического анализа.

4. Среды автоматизированного перевода текстов (CAT), такие как OmegaT[9] и отечественный онлайн-сервис Tolma.ch[10], выступающие альтернативой для таких коммерческих решений, как Trados Studio[11], и объединяющие в рамках единого программного решения такие функциональные возможности, как память переводов, подключение электронных словарей и глоссариев (в случае OmegaT в том числе электронных словарей, совместимых с открытым и свободным словарным агрегатором GoldenDict[12]). Сервис Tolma.ch также поддерживает командную работу над проектом в распределенном онлайн-формате.

Что касается непосредственно отечественных разработок, в каталоге российского ПО, рекомендуемого для импортозамещения, также есть ряд коммерческих решений. Они используются в качестве альтернативы зарубежному программному обеспечению (рубрика «Лингвистическое ПО»)[13]. Практический интерес, в частности, представляют российские платформы, на которых реализуются современные алгоритмы генеративного искусственного интеллекта и машинного обучения. Например, к такому ПО относится Naumen Erudite, развивающаяся платформа искусственного интеллекта для «создания голосовых роботов и чат-ботов и управления их работой»[14]. Она является собой альтернативу продуктам Google ASR/TTS, IBM Watson и т. п.

Совместимые с Linux отечественные программные продукты представлены в едином реестре российского ПО в рубрике «07. Лингвистическое программное обеспечение» и включают в себя парсеры и семантические анализаторы (07.01), средства речевого перевода (07.02), средства распознавания символов (07.03), средства распознавания и синтеза речи (07.04), средства автоматизированного перевода (07.05), электронные словари (07.06) и средства проверки правописания (07.07)[15].

Таким образом, модель цифрового рабочего процесса лингвиста в условиях перехода на операционные системы семейства Linux включает в себя: 1) собственно ОС на базе ядра Linux, что в условиях импортозамещения подразумевает одну из отечественных разработок, таких как ALT Linux (Альт СП), РЕД ОС и Astra Linux; 2) комплекс программного обеспечения общего назначения, разработанный на основе синтеза существующих открытых и свободных программных продуктов, совместимых с ОС Linux, а также отечественных коммерческих продуктов, разработанных в рамках программы импортозамещения; 3) комплекс лингвистического программного обеспечения для решения профессиональных задач, также представляющий собой синтез открытых международных и проприетарных отечественных решений в соответствии с потребностями каждой конкретной организации или научного коллектива.

ЗАКЛЮЧЕНИЕ

Результаты предпринятого исследования продемонстрировали высокую степень совместимости существующего лингвистического программного обеспечения с операционными системами

[1]URL: https://www.python.org/ (дата обращения: 23.05.2025).
[2]URL: https://www.nltk.org/ (дата обращения: 23.05.2025).
[3]URL: https://spacy.io/ (дата обращения: 23.05.2025).
[4]URL: https://textblob.readthedocs.io/en/dev/ (дата обращения: 23.05.2025).
[5]URL: https://textacy.readthedocs.io/en/latest/ (дата обращения: 23.05.2025).
[6]URL: https://code.visualstudio.com/ (дата обращения: 23.05.2025).
[7]URL: https://vscodium.com/ (дата обращения: 23.05.2025).
[8]URL:https://www.fon.hum.uva.nl/praat/ (дата обращения: 26.05.2025).
[9]URL:https://omegat.org/ (дата обращения: 26.05.2025).
[10]URL: https://tolma.ch/ (дата обращения: 26.05.2025).
[11]URL:https://www.trados.com/product/studio/ (дата обращения: 26.05.2025).
[12]URL: https://github.com/goldendict/goldendict (дата обращения: 26.05.2025).
[13]URL: https://catalog.arppsoft.ru/replacement/section_6116601 (дата обращения: 27.05.2025).
[14]URL: https://catalog.arppsoft.ru/product/6057858 (дата обращения: 27.05.2025).
[15]URL: https://reestr.digital.gov.ru/ (дата обращения: 27.05.2025).

семейства Linux, а также в полной мере подтвердили возможность организации полноценного цифрового рабочего процесса лингвиста, построенного на базе одной из операционных систем на базе ядра Linux и совместимого ПО общего и профессионального назначения, в том числе отечественного. Такой цифровой рабочий процесс представляется независимым от иностранных проприетарных решений и при этом характеризуется обширным функциональным потенциалом для решения широкого круга задач в области теоретической и прикладной лингвистики, связанных с автоматизированной обработкой и анализом текста и звучащей речи.

Существенными свойствами предложенной модели цифрового рабочего процесса лингвиста выступают:

1) способность гибкого синтеза свободного ПО с открытым кодом и коммерческих разработок;
2) высокая совместимость используемых в рамках данного рабочего процесса программных продуктов с проприетарными решениями, типичными для ОС Microsoft Windows, в части используемых форматов файлов и реализуемых алгоритмов автоматизированной обработки языковых и речевых данных.

СПИСОК ИСТОЧНИКОВ

1. Антопольский А. Б. Цифровые лингвистические информационные ресурсы. Определение объекта и каталогизация // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2021. № 3. С. 27–36.
2. Каменский М. В. Информационно-технологическое обеспечение оптимизации научно-исследовательской деятельности по теоретической и прикладной лингвистике в условиях цифровизации // Гуманитарные и юридические исследования. 2021. № 4. С. 208–218.
3. Горожанов А. И. Архитектура сбалансированного лингвистического корпуса, полученного автоматическим путем (опыт Московского государственного лингвистического университета) // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2024. Вып. 11 (892). С. 24–30.
4. Гуртуева И. А. Корпусное исследование акцентной русской речи // Человек - язык - компьютер. Исследователи будущего: Материалы Научно-практической (заочной) конференции с международным участием. Москва, 25 декабря 2023 года /отв. ред. А. И. Горожанов; редколлегия: А. А. Альварес Соллер, Д. В. Степанова, Л. А. Фурсова и др. М.: ФГБОУ ВО МГЛУ, 2024. С. 56–62.
5. Gorozhanov, A. I., Guseynova, I. A. Korpusanalyse der Konstituenten Grammatischer Kategorien im Literarischen Text mit Berücksichtigung der Linguoregionalen Komponente // Журнал Сибирского федерального университета. Серия: Гуманитарные науки. 2020. Т. 13. № 12. С. 2035 – 2048. DOI 10.17516/1997-1370-0702.
6. Аветисян А. И. Искусственный интеллект в гуманитарной сфере. Угрозы и возможности // Вестник Российской академии наук. 2024. Т. 94. № 7. С. 623–628.
7. Авраменко А. П. Большие языковые модели в лингвистике и лингводидактике: монография / под ред. А. Э. Левицкого, В. В. Терновского, В. А. Фадеева. М.: КДУ, Добросвет, 2023.
8. Поликарпов А. М. Филологическое знание в цифровой цивилизации // Сборник тезисов по итогам Профессорского форума 2019 «Наука. Образование. Регионы» / отв. ред. А.А. Громский, гл. ред. В. В. Гриб, председатель редсовета В. М. Филиппов. М.: Общероссийская общественная организация «Российское профессорское собрание», 2019. Т. 1. С. 182 – 184.
9. Шейко А. М. Digital Humanities и цифровая филология: истоки и перспективы развития // Теория языка и межкультурная коммуникация. 2023. № 3 (50). С. 259 – 273.
10. Гималетдинова Г. К., Довтаева Э. Х. Сентимент-анализ читательского комментария: автоматизированная VS ручная обработка текста // Ученые записки Казанского университета. Серия: Гуманитарные науки. 2021. Т. 163. № 1. С. 65 – 80.
11. Максименко О. И. Автоматизированный дистрибутивно-статистический анализ как системная обработка текста // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2019. Т. 10. № 1. С. 92 – 100.
12. Бойчук Е. И. Автоматизированный анализ ритма рекламного текста // Верхневолжский филологический вестник. 2021. № 1 (24). С. 137 – 144.
13. Фаткулин Б. Г. Прикладная лингвистика на службе китаеведения: автоматизация загрузки контента на заданную тему из китайской энциклопедии «БАЙДУ БАЙКЕ» с помощью специального программного обеспечения в рамках операционной системы LINUX // Россия и Китай: история и перспективы сотрудничества : Материалы V международной научно-практической конференции, Благовещенск-Хэйхэ-Харбин, 18 – 23 мая 2015

года / отв. ред.: Д. В. Буяров, Д. В. Кузнецов, Н. В. Киреева. Благовещенск – Хэйхэ – Харбин: Благовещенский государственный педагогический университет, 2015. Т. 5. С. 374–378.

REFERENCES

1. Antopolsky, A. B. (2021). Digital linguistic information resources. The definition of the object and cataloging. Automatic documentation and mathematical linguistics, 3, 27–36. (In Russ.)
2. Kamensky, M. V. (2021). Information technologies in optimizing scientific research in the sphere of theoretical and applied linguistics in the digital age. Humanities and law research, 4, 208–218. (In Russ.)
3. Gorozhanov, A. I. (2024). Architecture of a balanced linguistic corpus built automatically (experience of Moscow State Linguistic University). Vestnik of Moscow State Linguistic University. Humanities, 11(892), 24–30. (In Russ.)
4. Gurtueva, I. A. (2024). Corpus study of Russian accent speech. Human – language – computer. Issledovateli budushchego (pp. 56–62): Materialy Nauchno-prakticheskoj (zaochnoj) konferencii s mezhdunarodnym uchastiem. Moscow, 2023, December 25. Moscow: Moscow State Linguistic University. (In Russ.)
5. Gorozhanov, A. I., Guseynova, I. A. (2020). Korpusanalyse der Konstituenten Grammatischer Kategorien im Literarischen Text mit Berücksichtigung der Linguoregionalen Komponente. Journal of Siberian Federal University. Humanities & Social Sciences, 13(12), 2035–2048. DOI 10.17516/1997-1370-0702.
6. Avetisyan, A. I. (2024). Artificial intelligence in the humanitarian field. Threats and opportunities. Herald of the Russian Academy of Sciences, 94(7), 623–628. (In Russ.)
7. Avramenko, A. P. (2023). Bol'shie yazykovye modeli v lingvistike i lingvodidaktike = Large Language Models in Linguistics and Linguodidactics. Moscow: KDU, Dobrosvet. (In Russ.)
8. Polikarpov, A. M. (2019). Filologicheskoe znanie v tsifrovoi tsivilizatsii = Philological knowledge in digital civilization. In. Gromskij, A. A. (Ed.), Sbornik tezisov po itogam Professorskogo foruma 2019 «Nauka. Obrazovanie. RegionY» (vol. 1, pp. 182–184): collection of papers. Moscow: Obshcherossijskaya obshchestvennaya organizaciya "Rossijskoe professorskoe sobranie." (In Russ.)
9. Sheiko, A. M. (2023). Digital humanities and digital philology: origins and future. Teoriya yazyka i mezhkul'turnaya kommunikatsiya, 3(50), 259–273. (In Russ.)
10. Gimaletdinova, G. K., Dovtaeva, E. Kh. (2021). Sentiment analysis of reader comments: automated vs manual text processing. Uchenye zapiski kazanskogo universiteta. Seriya: gumanitarnye nauki, 163(1), 65–80. (In Russ.)
11. Maksimenko, O. I. (2019). Automatic Distributive-Statistic Analysis As System Text Processing. RUDN journal of language studies, semiotics and semantics, 10(1), 92–100. (In Russ.)
12. Boichuk, E. I. (2021). Automated analysis of the rhythm of advertising text. Verhnevolzhski philological bulletin, 1(24), 137–144. (In Russ.)
13. Fatkulin, B. G. (2015). Applied linguistics at the service of the Russian sinology: automate content downloading from the Chinese encyclopedia BAIDU BAIKE using the OS LINUX opportunities. Rossiya i Kitai: istoriya i perspektivy sotrudnichestva (vol. 5, pp. 374 –378): Proceedings of the 5th International Scientific and practical conference, 2015, May 18–23. Blagoveshchensk – Kheikhe – Kharbin. (In Russ.)

ИНФОРМАЦИЯ ОБ АВТОРЕ

Каменский Михаил Васильевич

доктор филологических наук, доцент
профессор департамента лингвистики факультета международных отношений
Северо-Кавказского федерального университета

INFORMATION ABOUT THE AUTHOR

Kamensky Mikhail Vasilyevich

Doctor of Philology (Dr. habil.), Associate Professor
Professor at the Department of Linguistics
Faculty of International Relationships, North-Caucasus Federal University

Статья поступила в редакцию	10.07.2025	The article was submitted
одобрена после рецензирования	12.08.2025	approved after reviewing
принята к публикации	15.09.2025	accepted for publication